EDA on Hotel booking analysis

# Introduction -

We have got the hotel bookings dataset. Our motto is to perform Exploratory Data Analysis on the given hotel booking CSV data set & to get a comfortable conclusion about general fluctuation in hotel booking and how things ruling in hotel bookings are related with each other.

# Dataset -

This data set contains booking information for a city hotel and resort hotel. It includes information such as type of the hotel, how many time it is cancelled, lead time, arrival date & month, stays in weekend nights, stays in week nights, children, adults, babies, country, market segment, distribution channel, is it repeated guest, previous cancellation, previous bookings, reserved room type, agent, company, customer type, required car parking & reservation status etc.

|  |  |  |
| --- | --- | --- |
|  | Understanding the whole dataset with help of column names. Given below is clearly mentioned.  hotel : Name of hotels  is\_canceled : Indicating the booking was cancelled (1) or not cancelled (0)  lead\_time : Number of days that elapsed between the entering data of booking  arrival\_date\_year : Year of arrival date  arrival\_date\_month : Month of arrival date  arrival\_date\_week\_number : week number of year of arrival  arrival\_date\_day\_of\_month : Day of arrival date  stays\_in\_week\_nights : The number of weekend nights ( saturday and sunday ) the guest stayed in hotel  stays\_in\_week\_nights : Number of week days ( monday to friday )the guest stayed in hotel  adults : Number of adults stayed in hotel  children : Number of children stayed in hotel  babies : Number of babies stayed in hotel  meal : type of meal booked by customers  country : country of origin  market\_segments : 'TA' means travel agent and 'TO' means team operators  distribution\_channel : Booking distribution channel  is\_repeated\_guest : Repeated guest (1) or not repeated guest (0)  previous\_cancellations : Number of booking that were cancelled by customers  previous\_bookings\_not\_canceled : Number of bookings that were not cancelled by customers  reserved\_room\_type : Code is represented by room whic is booked by customer  assigned\_room\_type : code is type of room assigned to the booking  booking\_changes : Number of charges made to the booking  deposit\_type : Indicates on the customer made a deposit to guarantee the booking  agent : ID for travel agency  company : Company ID entity that made booking or responsable for booking payment  days\_in\_waiting\_list : Number of days from booking to conformation booking  customer\_type : booking assuming for four categories  adr : Average daily rate sum of all loading transactions dividing by total number of staying nights  required\_car\_parking\_spaces : Car parking space required by customer  total\_of\_special\_requests : Total special requsts made by customer  reservation\_status : Reservation status, assuming in three categories  reservation\_status\_date : Date of the last status was set |  |

# Data cleaning –

1. detecting duplicate.
2. Handled null values

Null values column name company is dropped.

1. Null values in column name country were replaced by ‘others’.

Null values in columns names agent and children were replaced by ‘0’.

1. Detecting outliers

# Questions performed in EDA -

1. Which hotel is most preferred by customers?
2. Which month visitors visit highly?
3. Which type of room highly booked and preferred by customers?
4. Which year got a best sales?
5. Which hotels mostly cancelled by the customers?
6. Which type of customers highly visited on both hotels?
7. What is percentage of repeated guest?
8. What is the percentage distribution of deposit type?

# Libraries and tools used in EDA -

1. Pandas
2. Numpy
3. Seaborn
4. Matplotlib

# Graphs & plots been used -

1. Countplot
2. Pairplot
3. Heatmap
4. Boxplot
5. Distplot
6. Piechart

# Challenges faced –

1. Huge amount of data was present in dataset.
2. Dealt with some missing values.
3. Huge amount of null values were present in dataset.
4. Faced difficulties in understanding the data.

# Final outcome (Result) -

We learnt

1. Guest mostly preferred city hotel because city hotel has maximum bookings.
2. August is one of the month got high amount of visitors.
3. Code ‘A’ room are most preferred by customers because code ‘A’ room is highly booked by customers.
4. In 3 years of data we got to know sales of year 2016 are higher than year 2015 & 2017.
5. City hotel is mostly cancelled by the customers after booking.
6. Transient type of customer highly visited in both hotels.
7. 3.2% of customers are repeated guest.
8. 87.6% of data is distributed in deposit type.

# Conclusion –

We used the dataset that contains data about hotel bookings (which was already given). We solved the challenges that we faced while performing the project. It took us a lot of time to get the solution of questions that we performed in our colab notebook. After cleaning the data we simply grabbed. We had to understand the whole data first and then we performed the Exploratory Data Analysis to get the answer of our questions.